

# Self-control Dynamics for Sparsely Coded Networks with Synaptic Noise

D. Bollé and R. Heylen  
 Institute for Theoretical Physics  
 Katholieke Universiteit Leuven  
 B-3001, Leuven, Belgium  
 E-mail: desire.bolle@fys.kuleuven.ac.be

**Abstract**—For the retrieval dynamics of sparsely coded attractor associative memory models with synaptic noise the inclusion of a macroscopic time-dependent threshold is studied. It is shown that if the threshold is chosen appropriately as a function of the cross-talk noise and of the activity of the memorized patterns, adapting itself automatically in the course of the time evolution, an autonomous functioning of the model is guaranteed. This self-control mechanism considerably improves the quality of the fixed-point retrieval dynamics, in particular the storage capacity, the basins of attraction and the mutual information content.

## I. INTRODUCTION

Efficient neural network modelling requires an autonomous functioning independent from external constraints or control mechanisms. For fixed-point retrieval by an attractor associative memory model this requirement is mainly expressed by the robustness of its learning and retrieval capabilities against external noise, against malfunctioning of some of the connections and so on. Indeed, a model which embodies this robustness is able to perform as a content-addressable memory having large basins of attraction for the memorized patterns. Intuitively, one can imagine that these basins of attraction become smaller when the storage capacity gets larger. This might occur, e.g., in sparsely coded models (Okada, 1996 and references cited therein). Therefore, the necessity of a control of the activity of the neurons has been emphasized such that the latter stays the same as the activity of the memorized patterns during the recall process. This has led to several discussions imposing external constraints on the dynamics. However, the enforcement of such a constraint at every time step destroys part of the autonomous functioning of the network. To solve this problem, quite recently, a self-control mechanism has been introduced in the dynamics through the introduction of a time-dependent threshold in the transfer function (Dominguez & Bollé, 1998; Bollé, Dominguez & Amari 2000). This threshold is determined as a function of both the cross-talk noise and the activity of the memorized patterns in the network and adapts itself in the course of the time evolution.

Up to now only neural network models without synaptic noise have been considered in this context. The purpose of the present work is precisely to generalise this self-control mechanism when synaptic noise is allowed.

## II. THE MODEL

Let us consider a network of  $N$  binary neurons. At a discrete time step  $t$  the neurons  $\sigma_{i,t} \in \{0, 1\}$ ,  $i = 1, \dots, N$  are updated synchronously according to the rule

$$\sigma_{i,t+1} = F_{\theta_t, \beta}(h_{i,t}), \quad h_{i,t} = \sum_{j \neq i}^N J_{ij}(\sigma_{j,t} - a), \quad (1)$$

where  $J_{ij}$  are the synaptic couplings,  $a$  is the activity of the memorized patterns and  $h_{i,t}$  is usually called the “local field” of neuron  $i$  at time  $t$ . In general, the transfer function  $F_{\theta_t, \beta}$  can be a monotonic function with  $\theta_t$  a time-dependent threshold. Later on it will be chosen as

$$F_{\theta_t, \beta}(x) = \frac{1}{2}[1 + \tanh(\beta(x - \theta_t))]. \quad (2)$$

The “temperature”  $\beta = 1/T$  controls the thermal fluctuations, which are a measure for the synaptic noise (Hertz et al., 1991). In the sequel, for theoretical simplicity in the methods used, the number of neurons  $N$  will be taken to be sufficiently large.

The synaptic couplings  $J_{ij}$  themselves are determined by the covariance rule

$$J_{ij} = \frac{C_{ij}}{C\tilde{a}} \sum_{\mu=1}^p (\xi_i^\mu - a)(\xi_j^\mu - a), \quad \tilde{a} \equiv a(1 - a). \quad (3)$$

The memorized patterns  $\xi_i^\mu \in \{0, 1\}$ ,  $\mu = 1, \dots, p$  are independent identically distributed random variables (iidrv) with respect to  $i$  and  $\mu$  chosen according to the probability distribution

$$p(\xi_i^\mu) = a\delta(\xi_i^\mu - 1) + (1 - a)\delta(\xi_i^\mu). \quad (4)$$

The coefficients  $C_{ij} \in \{0, 1\}$  are iidrv with probability

$$\begin{aligned} \Pr\{C_{ij} = d\} &= [1 - (C/N)]\delta_{d,0} + (C/N)\delta_{d,1} \\ \Pr\{C_{ij} = C_{ji}\} &= (C/N)^2, \quad (C/N) \ll 1, \quad C > 0. \end{aligned} \quad (5)$$

This introduces the so-called extremely diluted asymmetric architecture with  $C$  measuring the average connectivity of the network (Derrida et al., 1987).

At this point we remark that the couplings (3) are of infinite range (each neuron interacts with infinitely many others) such that our model allows a so-called mean-field theory approximation. This essentially means that we focus on the dynamics of a single neuron while replacing all the other

neurons by an average background local field. In other words, no fluctuations of the other neurons are taken into account, not even in response to changing the state of the chosen neuron. In our case this approximation becomes exact because, crudely speaking,  $h_{i,t}$  is the sum of very many terms and a central limit theorem can be applied (Hertz et al., 1991).

It is standard knowledge by now that synchronous mean-field theory dynamics can be solved exactly for these diluted architectures (e.g., Bollé, 2004). Hence, the big advantage is that this will allow us to determine the precise effects from self-control in an exact way. We recall that the relevant parameters describing the solution of this dynamics are the retrieval overlap,  $m_t^\mu$ , between the memorized pattern,  $\xi_i^\mu$ , and the microscopic network state,  $\sigma_{i,t}$ , and the neural activity,  $q_t$ , given by, respectively

$$m_t^\mu \equiv \frac{1}{Na} \sum_i \xi_i^\mu \sigma_{i,t}, \quad q_t \equiv \frac{1}{N} \sum_i \sigma_{i,t}. \quad (6)$$

We remark that the  $m_t^\mu$  are normalized parameters within the interval  $[0, 1]$  which attain the maximal value 1 whenever the model succeeds in a perfect recall, i.e.,  $\sigma_{i,t} = \xi_i^\mu$  for all  $i$ .

In order to measure the retrieval quality of the recall process, we use the mutual information function (Bollé, Dominguez & Amari, 2000; Nadal, Brunel & Parga, 1998; Schultz & Treves, 1998 and references therein). In general, it measures the average amount of information that can be received by the user by observing the signal at the output of a channel (Blahut, 1990; Shannon, 1948). For the recall process of memorized patterns that we are discussing here, at each time step the process can be regarded as a channel with input  $\xi_i^\mu$  and output  $\sigma_{i,t}$  such that this mutual information function can be defined as (forgetting about the pattern index  $\mu$  and the time index  $t$ )

$$I(\sigma_i; \xi_i) = S(\sigma_i) - \langle S(\sigma_i | \xi_i) \rangle_{\xi_i}; \quad (7)$$

$$S(\sigma_i) \equiv - \sum_{\sigma_i} p(\sigma_i) \ln[p(\sigma_i)], \quad (8)$$

$$S(\sigma_i | \xi_i) \equiv - \sum_{\sigma_i} p(\sigma_i | \xi_i) \ln[p(\sigma_i | \xi_i)]. \quad (9)$$

Here  $S(\sigma_i)$  and  $S(\sigma_i | \xi_i)$  are the entropy and the conditional entropy of the output, respectively. These information entropies are peculiar to the probability distributions of the output. The term  $\langle S(\sigma_i | \xi_i) \rangle_{\xi_i}$  is also called the equivocation term in the recall process. The quantity  $p(\sigma_i)$  denotes the probability distribution for the neurons at time  $t$ , while  $p(\sigma_i | \xi_i)$  indicates the conditional probability that the  $i$ -th neuron is in a state  $\sigma_i$  at time  $t$ , given that the  $i$ -th pixel of the memorized pattern that is being retrieved is  $\xi_i$ . Hereby, we have assumed that the conditional probability of all the neurons factorizes, i.e.,  $p(\{\sigma_i\} | \{\xi_i\}) = \prod_i p(\sigma_i | \xi_i)$ , which is a consequence of the mean-field theory character of our model explained above. We remark that a similar factorization has also been used in Schwenker et al. (1996).

The calculation of the different terms in the expression (7) proceeds as follows. Formally writing  $\langle O \rangle \equiv \langle \langle O \rangle_{\sigma | \xi} \rangle_\xi = \sum_\xi p(\xi) \sum_\sigma p(\sigma | \xi) O$  for an arbitrary quantity  $O$  the conditional probability can be obtained in a rather straightforward

way by using the complete knowledge about the system:  $\langle \xi \rangle = a$ ,  $\langle \sigma \rangle = q$ ,  $\langle \sigma \xi \rangle = am$ ,  $\langle 1 \rangle = 1$ . The result reads (we forget about the index  $i$ )

$$\begin{aligned} p(\sigma | \xi) &= [\gamma_0 + (m - \gamma_0)\xi] \delta(\sigma - 1) \\ &\quad + [1 - \gamma_0 - (m - \gamma_0)\xi] \delta(\sigma), \\ \gamma_0 &= \frac{q - am}{1 - a} \end{aligned} \quad (10)$$

One can simply verify that this satisfies the averages

$$m = \frac{1}{a} \langle \langle \sigma \xi \rangle_{\sigma | \xi} \rangle_\xi \quad q = \langle \langle \sigma \rangle_{\sigma | \xi} \rangle_\xi \quad (11)$$

and those are precisely equal, for large  $N$ , to the parameters  $m$  and  $q$  mentioned above (Eq. (6)). Using the probability distribution of the patterns (Eq.(4)), we furthermore obtain

$$p(\sigma) \equiv \sum_\xi p(\xi) p(\sigma | \xi) = q\delta(\sigma - 1) + (1 - q)\delta(\sigma). \quad (12)$$

Hence the expressions for the entropies defined above become

$$S(\sigma) = -q \ln q - (1 - q) \ln(1 - q), \quad (13)$$

$$\begin{aligned} \langle S(\sigma | \xi) \rangle_\xi &= -a[m \ln(m) + (1 - m) \ln(1 - m)] \\ &\quad - (1 - a)[\gamma_0 \ln \gamma_0 + (1 - \gamma_0) \ln(1 - \gamma_0)]. \end{aligned} \quad (14)$$

Recalling eq. (7) this completes the calculation of the mutual information content of the present model.

### III. SELF-CONTROL DYNAMICS

It is standard knowledge (e.g., Derrida et al., 1987; Bollé, 2004) that the synchronous dynamics for diluted architectures can be solved exactly following the method based upon a signal-to-noise analysis of the local field (1) (e.g., Amari, 1977; Amari & Maginu, 1988; Okada, 1996; Bollé, 2004 and references therein). Without loss of generality we focus on the recall of one pattern, say  $\mu = 1$ , meaning that only  $m_t^1$  is macroscopic, i.e., of order 1 and the rest of the patterns causes a cross-talk noise at each time step of the dynamics. Supposing that the initial state of the network model,  $\{\sigma_{i,0}\}$ , is a collection of iidrv with mean zero and neural activity  $q_0$  and correlated only with memorized pattern 1 with an overlap  $m_0^1$ , then the full time evolution can be shown to be given by

$$m_{t+1}^1 = \langle F_{\theta_t, \beta}[(1 - a)M_t^1 + \omega_t] \rangle_\omega \quad (15)$$

$$q_{t+1} = am_{t+1}^1 + (1 - a)\langle F_{\theta_t, \beta}(-aM_t^1 + \omega_t) \rangle_\omega, \quad (16)$$

with

$$M_t^1 = \frac{m_t^1 - q_t}{1 - a}, \quad (17)$$

where we have averaged over the first pattern  $\xi^1$  and where the angular brackets indicate that we still have to average over the residual (cross-talk) noise  $\omega_t$  which can be written as

$$\omega_t = [\alpha Q_t]^{1/2} \mathcal{N}(0, 1), \quad Q_t = (1 - 2a)q_t + a^2 \quad (18)$$

with  $\mathcal{N}(0, 1)$  a Gaussian random variable with mean zero and variance unity and the (finite) loading defined by  $p = \alpha C$ .

Recalling the specific form of the transfer function (2) we explicitly have

$$\begin{aligned} & \langle F_{\theta_t, \beta}[-aM_t^1 + \omega_t] \rangle_\omega \\ &= \int_{-\infty}^{\infty} \frac{dy}{2\sqrt{2\pi\alpha Q_t}} e^{-y^2/\alpha Q_t} [1 + \tanh[\beta(-aM_t - \theta_t + y)]] \quad (19) \end{aligned}$$

and an analogous expression for  $\langle F_{\theta_t, \beta}[(1-a)M_t^1 + \omega_t] \rangle_\omega$ .

Of course, it is known that the quality of the recall process is influenced by the cross-talk noise at each time step of the dynamics. A novel idea is then to let the network itself autonomously counter this cross-talk noise at each time step by introducing an adaptive, hence time-dependent, threshold. This has been studied for neural network models at zero temperature, i.e., without synaptic noise where  $F_{\theta_t, \beta=\infty}(x) = \Theta(x - \theta_t)$ . For sparsely coded models, meaning that the pattern activity  $a$  is very small and tends to zero for  $N$  large, it has been found (Dominguez & Bollé, 1998; Bollé, Dominguez & Amari, 2000) that

$$\theta_t(a) = c(a)\sqrt{\alpha Q_t}, \quad c(a) = \sqrt{-2 \ln(a)} \quad (20)$$

makes the second term on the r.h.s of Eq.(16) asymptotically vanish faster than  $a$  such that  $q \sim a$ .

It turns out that the inclusion of this self-control threshold considerably improves the quality of the fixed-point retrieval dynamics, in particular the storage capacity, the basins of attraction and the information content. As an example we present in Fig. 1 the basin of attraction for the whole retrieval phase  $R$  for the self-control model with  $\theta_{sc}$  given by Eq. (20) and initial value  $q_0 = 0.01 = a$ , compared with a model where the threshold  $\theta_{opt}$  is selected for every loading  $\alpha$  by hand in an optimal way meaning that the information content  $i = \alpha I$  is maximized. The latter is non-trivial because it is even rather difficult, especially in the limit of sparse coding, to choose a threshold interval by hand such that  $i$  is non-zero. The basin of attraction is clearly enlarged with this self-control threshold choice and even near the border of critical storage the results are still improved. For more details we refer to Dominguez & Bollé (1998) and Bollé, Dominguez & Amari (2000). A similar threshold also works for sparsely coded sequential patterns (Kitani & Aoyagi, 1998) and even for non-sparse architectures as well (Bollé & Dominguez Carreta, 2000).

It is then worthwhile to examine whether such a self-control threshold can be found for networks with synaptic noise. No systematic study has been done in this case. The specific problem to be posed in analogy with the zero-temperature case is the following one. Can one determine a form for the threshold  $\theta_t$  in Eq. (19) such that the integral vanishes asymptotically faster than  $a$ ?

In contrast with the zero-temperature case, where due to the simple form of the transfer function, this threshold could be determined analytically (recall Eq. (20), a detailed study of the asymptotics of the integral in Eq. (19) gives no satisfactory analytic solution. Therefore, we have designed a systematic numerical procedure through the following steps:

- Choose a small value for the activity  $a'$ .

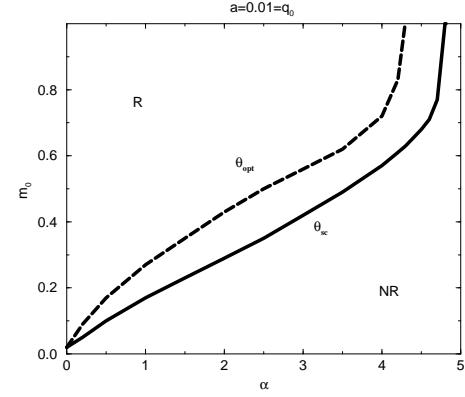


Fig. 1

THE BASIN OF ATTRACTION AS A FUNCTION OF  $\alpha$  FOR  $a = 0.01$  AND INITIAL  $q_0 = a$  FOR THE SELF-CONTROL MODEL (FULL LINE) AND THE OPTIMAL THRESHOLD MODEL (DASHED LINE) AT ZERO TEMPERATURE.

- Determine through numerical integration the threshold  $\theta'$  such that

$$\int_{-\infty}^{\infty} \frac{dx}{\sigma \sqrt{2\pi}} e^{-x^2/2\sigma^2} \Theta(x - \theta) \leq a' \quad \text{for } \theta > \theta' \quad (21)$$

for different values of the variance  $\sigma^2 = \alpha Q_t$ .

- Determine, as a function of the temperature  $T = 1/\beta$ , the value for  $\theta'_T$  such that

$$\int_{-\infty}^{\infty} \frac{dx}{2\sigma \sqrt{2\pi}} e^{-x^2/\sigma^2} [1 + \tanh[\beta(x - \theta)]] \leq a' \quad \text{for } \theta > \theta' + \theta'_T. \quad (22)$$

The second step leads, as expected, precisely to a threshold having the zero-temperature form Eq. (20). The third step determining the temperature dependent part  $\theta'_T$  leads to the results shown in Fig. 2. Intuitively it is seen that  $\theta'_T$  behaves

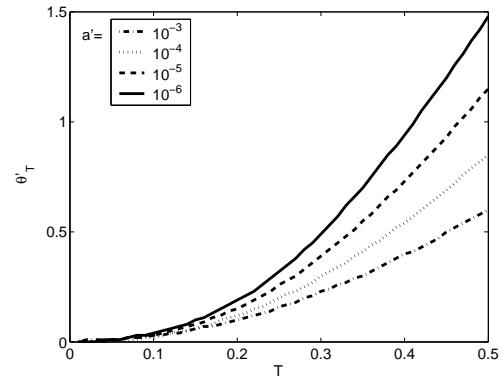


Fig. 2

THE TEMPERATURE DEPENDENT PART OF THE THRESHOLD  $\theta'_T$  AS A FUNCTION OF  $T$  FOR SEVERAL VALUES OF  $a'$

quadratically. Indeed, making a polynomial fit of these results we find that the linear term is negligible and that the quadratic term is of the form  $\theta'_T = -\frac{1}{2} \ln(a') T^2$ . Furthermore, the dependence of the coefficient of this quadratic term on the variance is very weak. Hence, we propose the following self-control threshold

$$\theta_t(a, T) = \sqrt{-2 \ln(a) \alpha Q_t} - \frac{1}{2} \ln(a) T^2. \quad (23)$$

Together with Eqs.(15)-(16) this relation describes the self-control dynamics of the network model with synaptic noise. This dynamical threshold is again a macroscopic parameter, thus no average must be taken over the microscopic random variables at each time step  $t$ .

At this point we want to make two remarks. First, for a binary layered network (Bollé & Massolo, 2000) the inclusion of a threshold of the form (20), although not designed for non-zero temperatures, is shown to still improve the retrieval quality for low pattern activities and low temperatures, in comparison with an optimal threshold model analogous to the one mentioned above. Secondly, in a recent study of an extremely diluted three-state neural network (Dominguez et al., 2002) based on information theoretic and mean-field theory arguments, a self-control threshold with a linear temperature correction term with coefficient 1 has been mentioned without any further details. In that specific model this self-control threshold is shown to improve the retrieval quality for low temperatures but it is not specified how much of the improvement is really due to the linear correction itself.

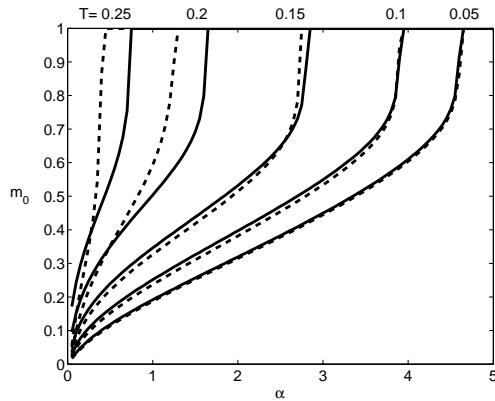


Fig. 3

THE BASIN OF ATTRACTION AS A FUNCTION OF  $\alpha$  FOR  $a = 0.01$  AND SEVERAL VALUES OF THE TEMPERATURE WITH (FULL LINES) AND WITHOUT (DASHED LINES) THE TEMPERATURE CORRECTION  $\theta'_T$  IN THE THRESHOLD.

We have solved this self-control dynamics, Eqs.(15)-(16) and Eq. (23), for our model with synaptic noise, in the limit of sparse coding, numerically. In particular, we have studied in detail the influence of the temperature dependent part of the threshold. Of course, we are only interested in the retrieval solutions with  $M > 0$  and carrying a non-zero information  $I$ .

We remark that all numerical calculations presented here are done for an appropriate number of time steps (at least of the order of a few hundred) in order to assure that a stable equilibrium point is reached.

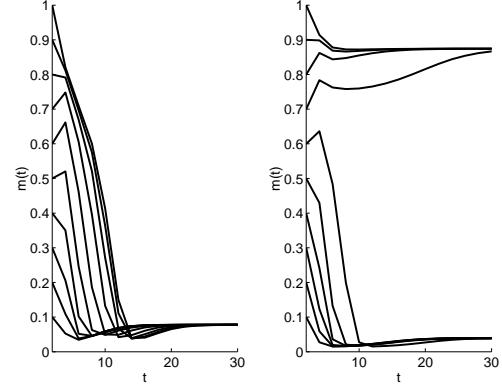


Fig. 4

THE EVOLUTION OF THE OVERLAP  $m_t$  FOR SEVERAL INITIAL VALUES  $m_0$  WITH  $a = 0.01$ ,  $T = 0.2$  AND  $\alpha = 1.5$  WITHOUT (LEFT) AND WITH (RIGHT) THE TEMPERATURE CORRECTION  $\theta'_T$  IN THE THRESHOLD.

The important features of the solution are illustrated in Figs. 3-5. In Fig. 3 we show the basin of attraction for the whole retrieval phase for the model with the temperature-zero threshold (20) (dashed curves) compared to the model with the temperature dependent threshold (23) (full curves) (compare also Fig. 1). We see that there is no clear improvement for low temperatures but there is a substantial one for higher temperatures. Even near the border of critical storage the results are still improved such that also the storage capacity itself is larger.

This is further illustrated in Fig. 4 where we compare the time evolution of the retrieval overlap  $m_t$  starting from several initial values,  $m_0$ , for the model with (right figure) and without (left figure) the quadratic temperature correction in the threshold. Here this temperature correction is absolutely crucial to force some of the overlap trajectories to go to the retrieval attractor  $m \approx 1$ . It really makes the difference between retrieval and non-retrieval in the model. At this point we remark that the influence of a linear temperature correction term has been examined also here but no real improvement has been found of the results for the temperature-zero threshold.

In Fig. 5 we plot the information content  $i$  as a function of the temperature for the self-control dynamics with the threshold (23) (full curves), respectively (20) (dashed curves). We see that, especially for small loading  $\alpha$  a substantial improvement of the information content is obtained.

#### IV. CONCLUSIONS

In this work we have generalized complete self-control in the dynamics of sparsely coded associative memory networks to models with synaptic noise. We have proposed an analytic form for the relevant macroscopic threshold consisting out of

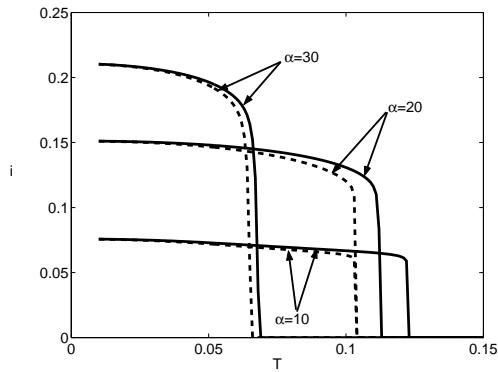


Fig. 5

THE INFORMATION CONTENT  $i$  AS A FUNCTION OF  $T$  FOR SEVERAL VALUES OF THE LOADING  $\alpha$  AND  $a = 0.001$  WITH (FULL LINES) AND WITHOUT (DASHED LINES) THE TEMPERATURE CORRECTION  $\theta'_T$  IN THE THRESHOLD.

the known form for temperature zero plus a quadratic temperature correction term dependent on the pattern activity. The consequences of this self-control mechanism on the quality of the recall process by the network have been studied.

We find that the basins of attraction of the retrieval solutions as well as the storage capacity are enlarged and that the mutual information content is maximized. This confirms the considerable improvement of the quality of recall by self-control, also for network models with synaptic noise.

This allows us to conjecture that this idea of self-control, allowing the network to function autonomously, might be relevant for other architectures in the presence of synaptic noise, and for dynamical systems in general, when trying to improve the basins of attraction and convergence times.

#### ACKNOWLEDGMENT

We are indebted to S. Goossens for some contributions at the initial stages of this work. One of the authors (DB) would like to thank D. Dominguez for stimulating discussions. This work has been supported by the Fund for Scientific Research-Flanders (Belgium).

#### REFERENCES

- [1] Amari S. (1977). Neural theory and association of concept information. *Biological Cybernetics*, **26**, 175-185.
- [2] Amari S. and Maginu K. (1988). Statistical neurodynamics of associative memory. *Neural Networks*, **1**, 63-73.
- [3] Blahut R.E. (1990). *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley.
- [4] Bollé D (2004). Multi-state neural networks based upon spin-glasses: a biased overview in *Advances in Condensed Matter and Statistical Mechanics* eds. Korutcheva E and Cuerno R., Nova Science Publishers, New-York, p. 321-349.
- [5] Bollé D. and Dominguez Carreta D. (2000). Mutual information and self-control of a fully connected low-activity neural network. *Physica A*, **286**, 401-416.
- [6] Bollé D., Dominguez D.R.C. and Amari S. (2000). Mutual information of sparsely coded associative memory with self-control and ternary neurons. *Neural Networks*, **13**, 455-462.
- [7] Bollé D. and Massolo G. (2000). Thresholds in layered neural networks with variable activity. *Journal of Physics A*, **33**, 2597-2609.
- [8] Derrida B., Gardner E., and Zippelius A. (1987). An exactly solvable asymmetric neural network model. *Europhysics Letters*, **4**, 167-173.
- [9] Dominguez D.R.C. and Bollé D. (1998). Self-control in sparsely coded networks. *Physical Review Letters*, **80**, 2961-2964.
- [10] Dominguez D.R.C., Korutcheva E., Theumann W.K., and Erichsen Jr. R. (2002). Flow diagrams of the quadratic neural network. *Lecture Notes in Computer Science*, **2415**, 129-134.
- [11] Hertz J., Krogh A. and Palmer R.G. (1991). *Introduction to the Theory of Neural Computation*. Addison-Wesley, Redwood City.
- [12] Kitano K. and Aoyagi T. (1998). Retrieval dynamics of neural networks for sparsely coded sequential patterns, *Journal of Physics A*, **31**, L613-L620.
- [13] Nadal J-P., Brunel N. and Parga N. (1998). Nonlinear feedforward networks with stochastic outputs: infomax implies redundancy reduction. *Network: Computation in Neural Systems*, **9**, 207-217.
- [14] Okada M. (1996). Notions of associative memory and sparse coding. *Neural Networks*, **9**, 1429-1458.
- [15] Schwenker F., Sommer F.T., and Palm G. (1996). Iterative retrieval of sparsely coded associative memory patterns. *Neural Networks*, **9**, 445-455.
- [16] Schultz S. and Treves A. (1998). Stability of the replica-symmetric solution for the information conveyed by a neural network. *Physical Review E*, **57**, 3302-3310.
- [17] Shannon C.E. (1948). A mathematical theory for communication. *Bell Systems Technical Journal*, **27**, 379.